

## *Journal of Applied Life Sciences International*

12(2): 1-12, 2017; Article no.JALSI.34058  
ISSN: 2394-1103

# **“Pollard-walk” Butterfly Survey does not Warrant the Equity of Sampling-completeness among Butterfly Families: A Case Study with Tropical Butterfly Fauna in Bhutan**

Jean Béguinot<sup>1\*</sup>

<sup>1</sup>Department of Biogéosciences, Université de Bourgogne, F 21000 – Dijon, France.

### **Author's contribution**

*The sole author designed, analyzed and interpreted and prepared the manuscript.*

### **Article Information**

DOI: 10.9734/JALSI/2017/34058

Editor(s):

(1) T. Selvamuthukumar, Department of Entomology, Faculty of Agriculture, Annamalai University, Tamil Nadu, India.

Reviewers:

(1) Manoel Fernando Demétrio, Universidade Federal da Grande Dourados, Brazil.

(2) Collins Ayine Nsor, Kwame Nkrumah University of Science and Technology, Ghana.

Complete Peer review History: <http://www.sciedomain.org/review-history/19595>

**Original Research Article**

**Received 11<sup>th</sup> May 2017**  
**Accepted 12<sup>th</sup> June 2017**  
**Published 17<sup>th</sup> June 2017**

## **ABSTRACT**

Complete or even sub-complete inventories of biodiversity often remain well out of reach, especially when dealing with speciose taxonomic groups, such as insects in general and butterfly fauna in particular. Moreover, it is even uncertain that similar completeness levels may be reached among taxonomic subsets within a same surveyed taxonomic group. For example, one can wonder whether sampling butterfly fauna according to the commonly implemented “Pollard walk” consistently ensure a similar recording efficiency (i.e. similar sampling completeness) according to the different butterfly families. To address this issue and quantify the possible unevenness of sampling completeness between five main butterfly families, we consider a “Pollard walk” butterfly survey of tropical butterfly fauna, carried out along Sankosh River (Bhutan) by Arun P. Singh. Reasonably accurate estimates of the true total species richness and the corresponding sampling completeness were derived, for each butterfly family, by implementing the recently developed procedure of “least-biased” extrapolation of the species accumulation curve. Whereby, a reliable evaluation of the level of sampling completeness is obtained for each butterfly family. For the inventory under consideration, a significant scatter of sampling completeness levels between the five families is highlighted: sampling completeness levels range from 65% for *Lycaenidae* to quasi exhaustivity for *Papilionidae*, with intermediate levels for the other three families.

\*Corresponding author: E-mail: [jean-beguिनot@orange.fr](mailto:jean-beguिनot@orange.fr);

These findings revealed that “Pollard walk” survey may lead to appreciable differences of sampling completeness level among the different butterfly families. Besides, this study provides a reliable estimate of the total species richness of butterfly fauna in the surveyed ecosystem, amounting to no less than 280 species concentrated in a rather reduced area along Sankosh River.

**Keywords:** Sankosh River; inventory; least biased extrapolation; species accumulation; estimation; species richness; species diversity.

## 1. INTRODUCTION

Incomplete inventories of biodiversity become increasingly frequent, as surveys progressively address taxonomic groups giving rise to highly species-rich assemblages, as is often the case, for example, with invertebrate faunas. Accordingly, most of published inventories of biodiversity are admittedly *more or less incomplete* [1], at least at the local/regional scales. High costs investments and shortage of available time to be devoted to these studies are arguably the main reasons for the current scarcity of quasi-exhaustive investigations of the kind. As a consequence, reliably assessing the actually reached level of sampling completeness is a major, *first issue* for most biodiversity surveys. Fortunately, it now becomes possible to address this issue properly, using reasonably accurate numerical extrapolations of the species accumulation process.

A *second issue*, also directly related to sampling incompleteness, is whether or not the achieved level of completeness substantially differ between taxonomical subsets within the sampled taxonomic set under consideration. For example, would substantial inequity exist or not between the levels of sampling completeness among butterfly families, after a partial inventory of butterfly fauna from a given surveyed ecosystem? This second issue, also, is of great importance since a fair equity between the levels of sampling completeness between taxonomical subsets is obviously a necessary condition to authorize reliable comparisons between the levels of species richness of each of these subsets respectively.

In case of substantial inequity in this respect, a *third issue* would be to reliably estimate, for each subset, the extent of additional sampling effort that should be required, so as to finally reach a similar level of sampling completeness among all taxonomic subsets involved.

Hereafter, we address these three successive questions, in the course of the extrapolative

analysis of the extensive survey of lowland forest butterflies of the Sankosh River Catchment (Bhutan) formerly published by Arun P. SINGH [2]. The derived expectations are challenged with those obtained using the empirical extrapolative model developed by Singh & Pandey [3] for the butterfly fauna of the Indian subcontinent. In particular, we try to highlight to what extent the Pollard Walk monitoring procedure actually allows (or not !) fairly even levels of sampling completeness among the different butterfly families (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperidae*), using the field data recorded by Arun P. SINGH.

## 2. MATERIALS AND METHODS

As just mentioned, Arun P. SINGH published a detailed assessment of part of the butterfly fauna in lowland forest in the vicinity of the Sankosh River Catchment (south-west Bhutan), as a part of a biodiversity assessment, prior to a planned hydroelectric power project in this area. The five butterfly families (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperidae*) that co-occur in the vicinity of the Sankosh River Catchment were surveyed simultaneously according to the usually implemented “Pollard-walk” procedure [4,5]. Other details regarding the sampling method and the recorded data are available in the *on line* publication of the author [2]. Accordingly, this information is not recalled here any longer. The achieved survey (1731 observed individuals) records the occurrence of 213 species of butterflies. This survey, yet, remains uncomplete, considering the substantial proportion of singletons (species only recorded once in the course of the survey). Accordingly, this survey, as most of the kind, is eligible to the *three questions* evoked in Introduction.

In this perspective, an accurate estimation of the number of still unrecorded species and reliable extrapolations of the species accumulation curves beyond the actually achieved sampling-size, were to be implemented: (i) for the surveyed butterfly group as a whole and (ii) for each of the represented families within.

## 2.1 Numerical Extrapolation of Species Accumulation beyond the Achieved Sampling Size

As sampling size increases, the number of recorded species is monotonically growing, at first rapidly and then less and less quickly. The so-called 'Species Accumulation Curve'  $R(N)$  accounts for the growth kinetics of the number of recorded species  $R$  with increasing sampling size  $N$  ( $N$ : typically, the number of observed individuals during sampling). The mathematical expression (and thus the details of the shape) of the Species Accumulation Curve are dependent upon both the total species richness of the sampled assemblage of species and the degree of heterogeneity of the species abundance distribution within the sampled assemblage of species [1]. This would apparently make the extrapolation of the Species Accumulation Curve rather difficult to compute, since both preceding factors are unknown *a priori*. Yet, the numbers  $f_1, f_2, f_3, f_4, \dots, f_x, \dots$  of those species recorded respectively 1-, 2-, 3-, 4-, ...,  $x$ - times during sampling are directly dependent also upon the total species richness *and* the degree of heterogeneity of the species abundances. This explains why these numbers  $f_1, f_2, f_3, f_4, \dots$  may serve as an appropriate basis from which to extrapolate the Species Accumulation Curve, beyond the actual size of the sample under consideration. In particular, the most commonly used estimators of the number of unrecorded species (i.e. *non-parametric* estimators such as 'Chao' and the series of 'Jackknife') are computed from the recorded values of the first numbers  $f_x$  [6]. In practice, a problem remains however: as already mentioned, each of these different types of estimators provides a substantially distinct estimate and none among these estimators remains consistently the more appropriate. Accordingly the traditional practice has become to consider together all of them without making any choice [7], an admittedly frustrating situation!

Yet, it has been shown recently that although none of the available estimators consistently remains the more accurate [8], each of them may prove, in turn, being the less biased, depending on the value taken by  $f_1$  as compared to the other  $f_{x>1}$  [9]. Accordingly, in practice, the most appropriate – i.e. *the least biased* – estimator of the number of unrecorded species may be selected by comparing the value of  $f_1$  to the values of the other  $f_x$  for  $x > 1$  [9,10]. Selecting this way the least-biased type of estimator

thereby provides the best possible estimate of the number  $\Delta$  of "missing" species and, in turn, the best estimate of the total species richness  $S_t$  of the partially sampled assemblage. In addition, the less biased expression for the *extrapolation* of the species accumulation curve  $R(N)$  is straightforwardly derived.

In practice, the formulations summarised in Appendix 1 provide (i) the expressions of  $\Delta$ ,  $S_t$  and  $R(N)$ , according to each of the most commonly used types of nonparametric estimators and (ii) the *key to select*, among them, the less biased estimator and, thereby, the less-biased expressions for  $\Delta$ ,  $S_t$  and  $R(N)$ . Also, in order to reduce the influence of drawing stochasticity, which affects the *as-recorded* values of the  $f_x$ , it is advisable to regress the as-recorded distribution of the numbers  $f_x$  versus  $x$ .

## 3. RESULTS

### 3.1 Estimations of Total Species Richness and Completeness per Family

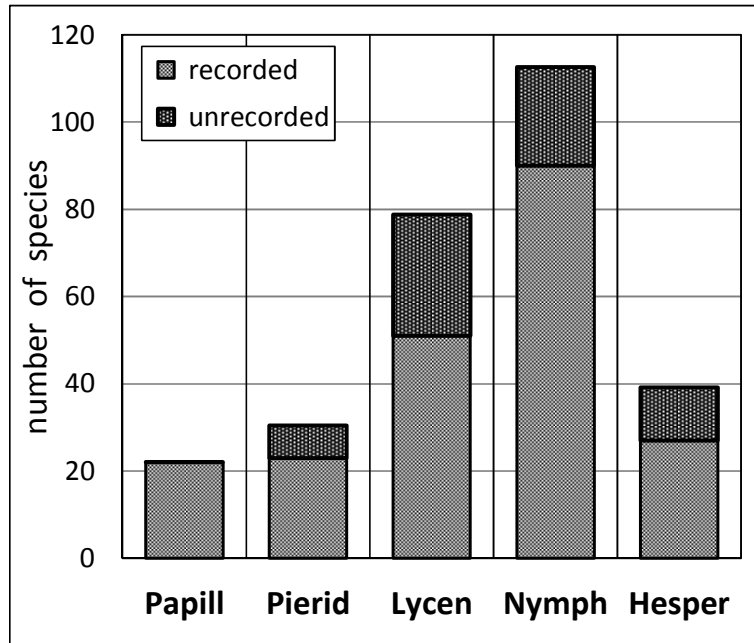
Table 1 details, for each butterfly family (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*) and for all these families together: the number of recorded species  $R_0$ , the selected least-biased type of nonparametric estimator, the estimated number  $\Delta$  of still unrecorded species, the resulting estimates of the "true" total species richness  $S_t$ . From these estimations of true species richness  $S_t$  are subsequently derived: the estimated level of sampling completeness  $R_0/S_t$  and the respective contributions (%) of each family to the overall butterfly species richness. In turn, Fig. 1. provides a graphical comparison of the number of recorded species, the estimated number of unrecorded species and the resulting estimation of true species richness for each of the five families of butterfly *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*.

The levels of sampling completeness of each of the five butterfly families, were immediately derived from these estimations (Figs. 2 and 3).

The respective contributions (%) of each family to the overall butterfly species richness (Table 1) were compared to the results of the Singh & Pandey empirical model [3] for North-East India (including Bhutan) and for two regions close to Bhutan (Sikkim and Darjeeling districts) (Table 2 and Fig. 4).

**Table 1.** For each butterfly family (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*) and all families together: the number of recorded species  $R_0$ , the selected least-biased type of nonparametric estimator, the estimated number  $\Delta$  of unrecorded species, the resulting estimates of the “true” total species richness  $S_t$ , the estimated level of sampling completeness  $R_0/S_t$ , the respective contributions (%) of each family to the overall butterfly species richness, according to the estimated true species richness  $S_t$  of each family. NB: the estimated species richness for all five families together (281 species) slightly differs from the sum of the estimated species richness of each of the five families (283), due to the fact that the involved estimators, although selected as least-biased, are not entirely unbiased however

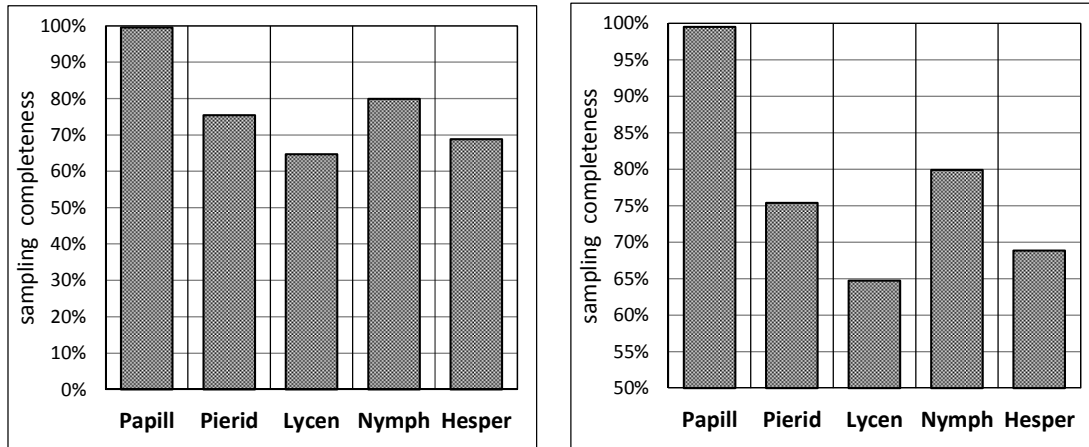
	Papill	Pierid	Lycen	Nymph	Hesper	all Fam
* nb. recorded sp. $R_0$	22	23	51	90	27	213
* least-biased estim.	Chao	JK.5	JK.4	JK.2	JK.2	JK.2
* nb. unrecorded sp. $\Delta$	0.1	7.5	27.8	22.6	12.2	68
* total sp. richness $S_t$	<b>22.1</b>	<b>30.5</b>	<b>78.8</b>	<b>112.6</b>	<b>39.2</b>	<b>281</b>
* sampl. completeness $R_0/S_t$	99 %	75 %	65 %	80 %	69 %	76 %
* contribution per family	7.8 %	10.8 %	27.8 %	39.8 %	13.8 %	100 %



**Fig. 1.** Histogram of the number of recorded species, the estimated number of unrecorded and, accordingly the estimated species richness of each of the five families of butterfly (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*) surveyed by A. P. SINGH at Sankosh River Catchment, Bhutan [2]. Derived from Table 1

Table 3 and Fig. 5 showed a comparison between crude field data (recorded numbers of species per family) and the estimated total species richness per family obtained either (i) by selecting the “least-biased” nonparametric estimator for each family [9] or (ii) by using the empirical model proposed by Singh & Pandey [3] for north-east India. Clearly, *crude field data* provides

unreliable appreciations of the true respective contributions of each family to the butterfly species richness and this, not only in absolute terms, but also in relative terms. On the other hand, the *empirical model* already provided significant progress, as it better approaches the values provided by the selected least-biased estimates taken as the *best available data source*.



**Figs. 2 and 3.** Estimated levels of sampling-completeness for the five families of butterfly (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*) simultaneously surveyed using Pollard-walk procedure. Fig. 3. is a zoom of Fig. 2, with ordinate beginning at 50%

**Table 2.** The respective contributions (%) to the overall butterfly species richness of each butterfly family (*Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*): (i) according to the least-biased extrapolation procedure for Sankosh River Catchment, Bhutan (this study); (ii) according to the expectations of Singh & Pandey model [3] for North-East India as a whole (including Bhutan) and for two regions close to Bhutan: Sikkim and Darjeeling district

Contributions per family to overall species richness	Papill	Pierid	Lycen	Nymph	Hesper
* Least-biased extrapolation for Sankosh river	7.8 %	10.8 %	27.8 %	39.8 %	13.8 %
* North-East India	7.3 %	6.1 %	30.2 %	34.2 %	22.2 %
* Sikkim	8.0 %	7.4 %	23.5 %	38.1 %	23.0 %
* Darjeeling district	11.1 %	12.2 %	18.3 %	48.1 %	10.3 %

**Table 3.** The number of recorded species and the estimated true species richness according to (i) the least-biased procedure (this study) and (ii) the model by Singh & Pandey [3], for each butterfly family, *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae* and for the five families together. NB: the estimated species richness of all five families together (281 species) slightly differs from the sum of the estimated species richness of each of the five families (283), due to the fact that the involved estimators, although selected as least-biased, are not entirely unbiased however

	Recorded number of species	Estimated species richness (this study)	Singh & Pandey model
<i>Papilionidae</i>	22	22,1	22,0
<i>Pieridae</i>	23	30,5	18,4
<i>Lycaenidae</i>	51	78,8	91,0
<i>Nymphalidae</i>	90	112,6	103,1
<i>Hesperiidae</i>	27	39,2	66,9
All families together	213	281	301,4

#### 4. DISCUSSION

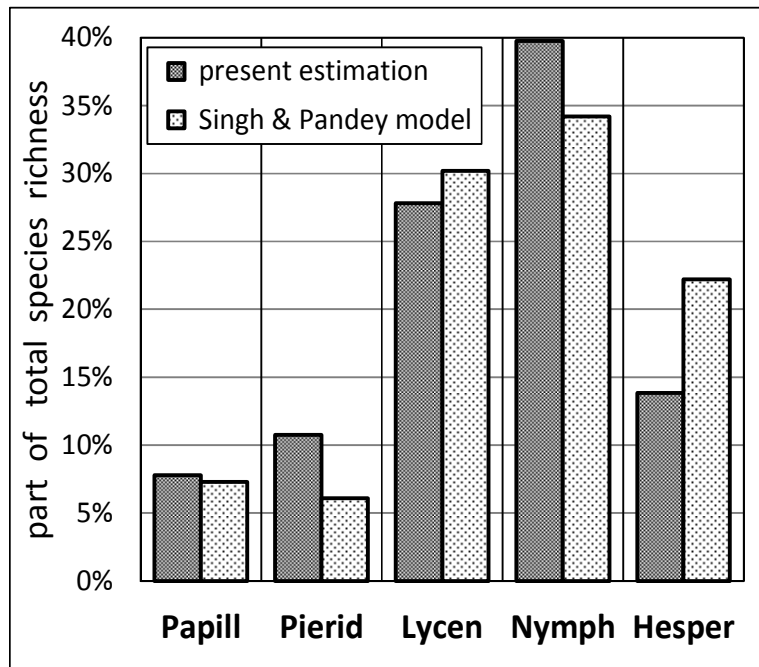
“Pollard walk” procedure, implemented for monitoring species abundances and carrying

surveys of butterfly fauna, has gained large recognition for decades. Yet, as probably any kind of sampling procedure, the Pollard walk method is not immune to some kind of criticism,

in particular regarding the uneven degree of detectability of butterfly species encountered along transect walks (see for example [11]). In this respect, the aim of the present case study is no more than trying to quantify, as accurately as possible, the expected consequence of the limitation above, in term of unevenness of sampling completeness among butterfly families after a Pollard walk inventory. Besides, the objective of this contribution was not to discuss the reasons at the origin of the uneven levels of completeness highlighted here, but, more practically, to draw attention to the *quantitative* consequences of unequal degrees of representativeness of butterfly inventories, according to each surveyed families.

Reliably estimating sampling completeness per family implies, first, to extrapolate appropriately the process of species accumulation so as to estimate, as accurately as possible, the true species richness. This was conducted using the recently developed procedure of selection of the “least-biased” nonparametric estimator of the

number of still unrecorded species. The computed levels of sampling completeness were in partial agreement with the empirical expectations from the Singh & Pandey model [3] for North-East India. Yet, some discrepancy occurs with regard to the two families *Pieridae* and *Hesperiidae*, which were substantially underestimated and overestimated, respectively, by the empirical model, when compared to the least-biased extrapolations. However, globally, the model is hereby comforted and may serve as an approximate convenient surrogate to extrapolations, for the Indian subcontinent. The worst solution, indeed, would be to rely only upon crude recorded data, which confirmed as being rather unreliable (Table 3 and Fig. 5). Incidentally, the Singh & Pandey model emphasises that species within the family *Papilionidae* are especially easy to observe, identify and sample [3] (p. 85), which complies with the exceptionally high level of exhaustiveness computed here for the inventory of this family.



**Fig. 4.** The proportions (in term of estimated numbers of species per family) of the five butterfly families, *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*: (i) on the basis of least-biased nonparametric estimator [this study] and (ii) using the model by Singh & Pandey [3] for north-east India. Both approaches are in fairly good accordance with, yet, two exceptions for the proportions relative to the two families *Pieridae* and *Hesperiidae*, which, respectively, are substantially underestimated and overestimated by the model, when compared to the least-biased extrapolations. Indeed Singh & Pandey already pointed out the larger variances affecting the proportions of both families according to localities ([3] p. 88)

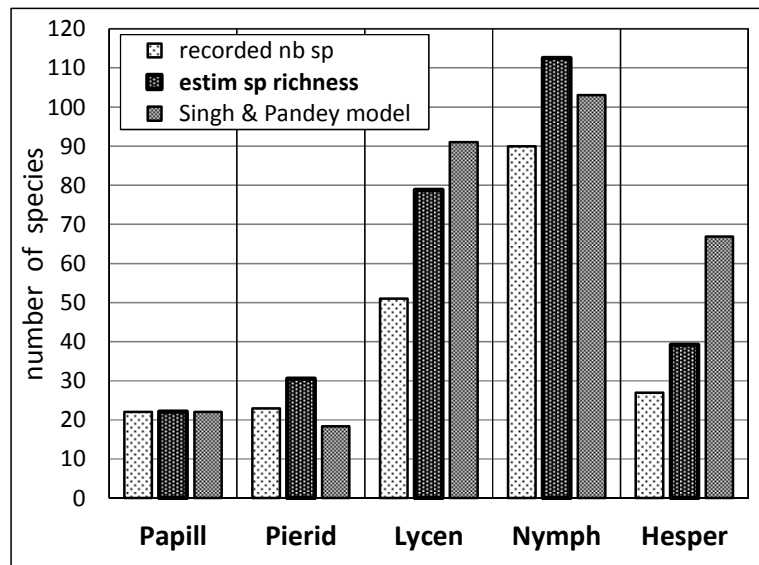


Fig. 5. The number of recorded species and the estimated true species richness according to (i) the least-biased procedure (this study) and (ii) the model by Singh & Pandey [3], for each butterfly family, *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*

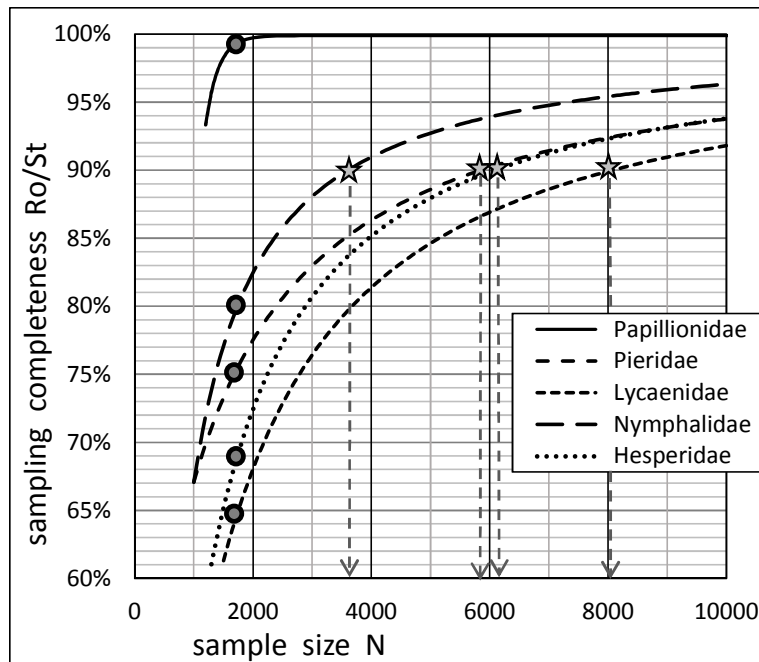


Fig. 6. Least-biased extrapolations of the Species Accumulation Curves (beyond the actually achieved sampling-size  $N_0 = 1731$  individuals: grey points) for the five butterfly families. Here, these extrapolations are effectively implemented to predict the additional sampling efforts required in order to ensure a 90% level of sampling completeness for each of the four under-sampled families *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*, respectively

However, being able to quantify the unevenness of sampling-completeness levels among butterfly families and, hence, to correct accordingly the recorded data in terms of relative contributions of

each families to the overall butterfly species richness of the surveyed area may well be considered as a first progress only. Continuing the inventory further, so that the less well-sampled families finally attain a same desirable high level of completeness would be a desirable *second step* of progress desirable to be reached. This implies, however, to plan the additional sampling efforts that would be respectively required for each family. The respective levels of additional efforts may be reliably predicted only by extrapolating the species accumulation curve for each family specifically. This, in turn, is made possible using the procedure of selection of the “least-biased” extrapolation ([9,10]; Appendix 1) already implemented for true species richness estimations. A convenient planning of the additional sampling efforts, differentiated according to each family, may then be derived accordingly.

Fig. 6 provides, for each family, the corresponding extrapolation and, thereby, the expected sampling effort required to reach any given level of inventory completeness. For example, aiming to reach a 90% sampling completeness would need to add  $\approx$  2000, 4200, 4500 and 6400 butterflies observations, considering families *Nymphalidae*, *Pieridae*, *Hesperiidae* and *Lycaenidae* respectively.

At last, coming back to the first aim of the survey conducted by A. P. SINGH - that is the butterfly species richness at the Sankosh River Catchment (Bhutan) - the least-biased estimates (Table 1) are: 22, 30, 79, 113, 39, 281 species respectively for the families *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae* and for the five families together.

## 5. CONCLUSION

Whether or not the commonly implemented “Pollard walk” method ensures fairly even degrees of sampling completeness according to the different families of butterfly was a question which, until then, remained pending. The recently developed “least-biased” procedure suggested that the “Pollard walk” method may actually lead to substantially unequal levels of sampling completeness among the different butterfly families. This, at least, was demonstrated above for the (partial) survey of the butterfly fauna at Sankosh River catchment (Bhutan) under consideration. The estimated levels of sampling completeness ranged from 65% (*Lycaenidae*) to 99% (*Papilionidae*), with families *Hesperiidae*,

*Pieridae*, *Nymphalidae* respectively sampled at 69%, 75% and 80% completeness levels.

Besides, this study has provided the opportunity to estimate at best the overall species richness of butterfly fauna around Sankosh River Catchment, which is estimated, at least, to 280 species, out of which 22, 30, 79, 113, 39 species belong respectively to the families *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperiidae*.

## ACKNOWLEDGEMENTS

Arun P. Singh is gratefully acknowledged for the achievement and detailed publication (SINGH 2012) of the inventories that serve as the data base for the present analysis.

## COMPETING INTERESTS

Author has declared that no competing interests exist.

## REFERENCES

1. Rajakaruna H, Drake DAR, Chan FT, Bailey SA. Optimizing performance of nonparametric species richness estimators under constrained sampling. *Ecology and Evolution*. 2016;6:7311-7322.
2. Singh AP. Lowland forest butterflies of the Sankosh River catchment, Bhutan. *Journal of Threatened Taxa*. 2012;4(12):3085-3102. (Available on line)
3. Singh AP, Pandey R. A model for estimating butterfly species richness of areas across the Indian subcontinent: species proportion of family *Papilionidae* as an indicator. *Journal of the Bombay Natural History Society*. 2004;101(1):79-89.
4. Pollard E. A method for assessing changes in the abundance of butterflies. *Biological Conservation*. 1977;12:115-134.
5. Pollard E. Monitoring butterfly abundance in relation to the management of a nature reserve. *Biological Conservation*. 1982; 24:317-328.
6. Gotelli NJ, Chao A. Measuring and estimating species richness, species diversity, and biotic similarity from sampling data. In: Levin SA (ed.) *Encyclopedia of Biodiversity*. Academic Press, Second Edition. 2013;5:195-211.
7. Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. *Philosophical Transactions*



- of the Royal Society London B. 1994;345:101-118.
8. Brose U, Martinez ND, Williams RJ. Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. Ecology. 2003;84(9):2364-2377.
9. Béguinot J. Theoretical derivation of a bias-reduced expression for the extrapolation of the species accumulation curve and the associated estimation of total species richness. Advances in Research. 2016;7(3):1-16.  
DOI:10.9734/AIR/2016/26387  
<hal-01367803>
10. Béguinot J. Extrapolation of the species accumulation curve associated to "Chao" estimator of the number of unrecorded species: A mathematically consistent derivation. Annual Research & Review in Biology. 2016;1(4):1-19.  
DOI:10.9734/ARRB/2016/30522  
<hal 01477263 >
11. Pellet J, Bried JT, Parietti D, Gander A, Heer PO, Cherix D, Arlettaz RI. Monitoring butterfly abundance: Beyond Pollard Walks. PLoS One. 2012;7(7):e41396.
12. Béguinot J. An algebraic derivation of Chao's estimator of the number of species in a community highlights the condition allowing Chao to deliver centered estimates. ISRN Ecology. 2014;article ID 847328.  
DOI:10.1155/2014/847328  
<hal-1101415>
13. Béguinot J. When reasonably stop sampling? How to estimate the gain in newly recorded species according to the degree of supplementary sampling effort. Annual Research & Review in Biology. 2015;7(5):300-308.  
DOI :10.9734/ARRB/2015/18809  
<hal-01228695>
14. Béguinot J. Extrapolation of the species accumulation curve for incomplete species samplings: A new nonparametric approach to estimate the degree of sample completeness and decide when to stop sampling. Annual Research & Review in Biology. 2015;8(5):1-9.  
DOI:10.9734/ARRB/2015/22351  
<hal-01238720>
15. Béguinot J. On general mathematical constraints applying to the kinetics of species discovery during progressive sampling: Consequences on the theoretical expression of the species accumulation curve. Advances in Research. 2016;8(5):1-17.  
DOI:10.9734/AIR/2016/31791  
<hal-01516141>

## APPENDIX 1

### Bias-reduced extrapolation of the Species Accumulation Curve and associated bias-reduced estimation of the number of missing species, based on the recorded numbers of species occurring 1 to 5 times

Consider the survey of an assemblage of species of size  $N_0$  (with sampling effort  $N_0$  typically identified either to the number of recorded individuals or to the number of sampled sites, according to the inventory being in terms of either species abundances or species incidences), including  $R(N_0)$  species among which  $f_1, f_2, f_3, f_4, f_5$ , of them are recorded 1, 2, 3, 4, 5 times respectively. The following procedure, designed to select the less-biased solution, results from a general mathematical relationship that constrains the theoretical expression of *any* theoretical Species Accumulation Curves  $R(N)$  (see [9,12,13,14,15]):

$$\partial^x R(N)/\partial N^x = (-1)^{(x-1)} f_{x(N)} / C_{N,x} \approx (-1)^{(x-1)} (x!/N^x) f_{x(N)} \quad (\approx \text{as } N \gg x) \quad (\text{A.1})$$

Compliance with the mathematical constraint (equation (A.1)) warrants *reduced-bias* expression for the extrapolation of the Species Accumulation Curves  $R(N)$  (i.e. for  $N > N_0$ ). Below are provided, accordingly, the polynomial solutions  $R_x(N)$  that respectively satisfy the mathematical constraint [1], considering increasing orders  $x$  of derivation  $\partial^x R(N)/\partial N^x$ . Each solution  $R_x(N)$  is appropriate for a given range of values of  $f_1$  compared to the other numbers  $f_x$  (according to [9]):

$$* \text{ for } f_1 \text{ up to } f_2 \rightarrow R_1(N) = (R(N_0) + f_1) - f_1 \cdot N_0/N$$

$$* \text{ for } f_1 \text{ up to } 2f_2 - f_3 \rightarrow R_2(N) = (R(N_0) + 2f_1 - f_2) - (3f_1 - 2f_2) \cdot N_0/N - (f_2 - f_1) \cdot N_0^2/N^2$$

$$* \text{ for } f_1 \text{ up to } 3f_2 - 3f_3 + f_4 \rightarrow R_3(N) = (R(N_0) + 3f_1 - 3f_2 + f_3) - (6f_1 - 8f_2 + 3f_3) \cdot N_0/N - (-4f_1 + 7f_2 - 3f_3) \cdot N_0^2/N^2 - (f_1 - 2f_2 + f_3) \cdot N_0^3/N^3$$

$$* \text{ for } f_1 \text{ up to } 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow R_4(N) = (R(N_0) + 4f_1 - 6f_2 + 4f_3 - f_4) - (10f_1 - 20f_2 + 15f_3 - 4f_4) \cdot N_0/N - (-10f_1 + 25f_2 - 21f_3 + 6f_4) \cdot N_0^2/N^2 - (5f_1 - 14f_2 + 13f_3 - 4f_4) \cdot N_0^3/N^3 - (-f_1 + 3f_2 - 3f_3 + f_4) \cdot N_0^4/N^4$$

$$* \text{ for } f_1 \text{ larger than } 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow R_5(N) = (R(N_0) + 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5) - (15f_1 - 40f_2 + 45f_3 - 24f_4 + 5f_5) \cdot N_0/N - (-20f_1 + 65f_2 - 81f_3 + 46f_4 - 10f_5) \cdot N_0^2/N^2 - (15f_1 - 54f_2 + 73f_3 - 44f_4 + 10f_5) \cdot N_0^3/N^3 - (-6f_1 + 23f_2 - 33f_3 + 21f_4 - 5f_5) \cdot N_0^4/N^4 - (f_1 - 4f_2 + 6f_3 - 4f_4 + f_5) \cdot N_0^5/N^5$$

The associated non-parametric estimators of the number  $\Delta_j$  of missing species in the sample [with  $\Delta_j = R(N=\infty) - R(N_0)$ ] are derived immediately:

$$* 0.6 f_2 < f_1 \leq f_2 \rightarrow \Delta_{j1} = f_1 ; R_1(N)$$

$$* f_2 < f_1 \leq 2f_2 - f_3 \rightarrow \Delta_{j2} = 2f_1 - f_2 ; R_2(N)$$

$$* 2f_2 - f_3 < f_1 \leq 3f_2 - 3f_3 + f_4 \rightarrow \Delta_{j3} = 3f_1 - 3f_2 + f_3 ; R_3(N)$$

$$* 3f_2 - 3f_3 + f_4 < f_1 \leq 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{J4} = 4f_1 - 6f_2 + 4f_3 - f_4 ; R_4(N)$$

$$* f_1 > 4f_2 - 6f_3 + 4f_4 - f_5 \rightarrow \Delta_{J5} = 5f_1 - 10f_2 + 10f_3 - 5f_4 + f_5 ; R_5(N)$$

**N.B. 1:** As indicated above (and demonstrated in details in [9]), this series of inequalities define the ranges that are best appropriate, respectively, to the use of each of the five estimators, JK-1 to JK-5. That is the respective ranges within which each estimator will benefit of minimal bias for the predicted number of missing species.

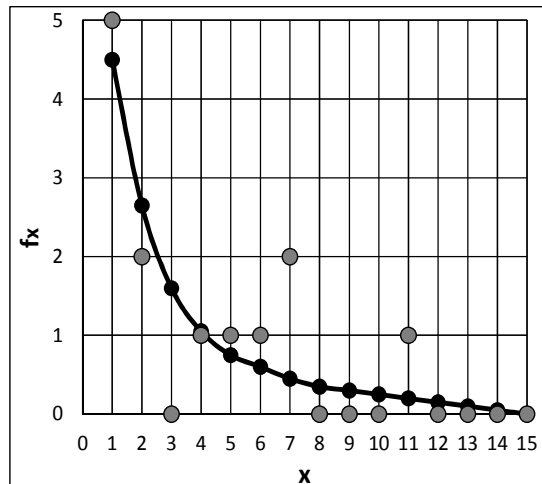
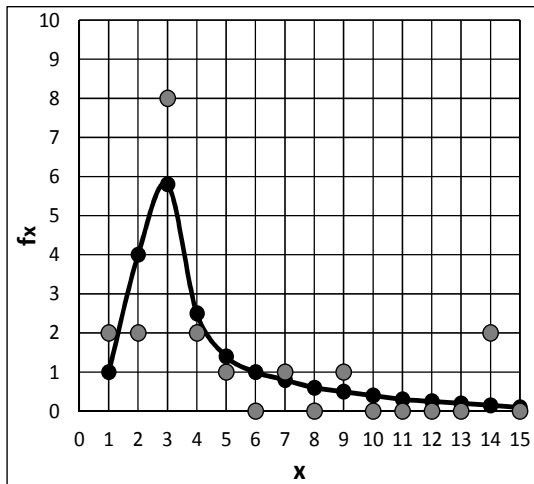
Besides, it is easy to verify that another consequence of these preferred ranges is that the selected estimator will *always* provide the highest estimate, as compared to the other estimators. Interestingly, this mathematical consequence, of general relevance, is in line with the already admitted opinion that all non-parametric estimators provide *under*-estimates of the true number of missing species [1,6]. Also, this shows that the approach initially proposed by Brose et al. [8] – which has regrettably suffered from its somewhat difficult implementation in practice – might be advantageously reconsidered, now, in light of the very simple selection key above, of *far much easier practical use*.

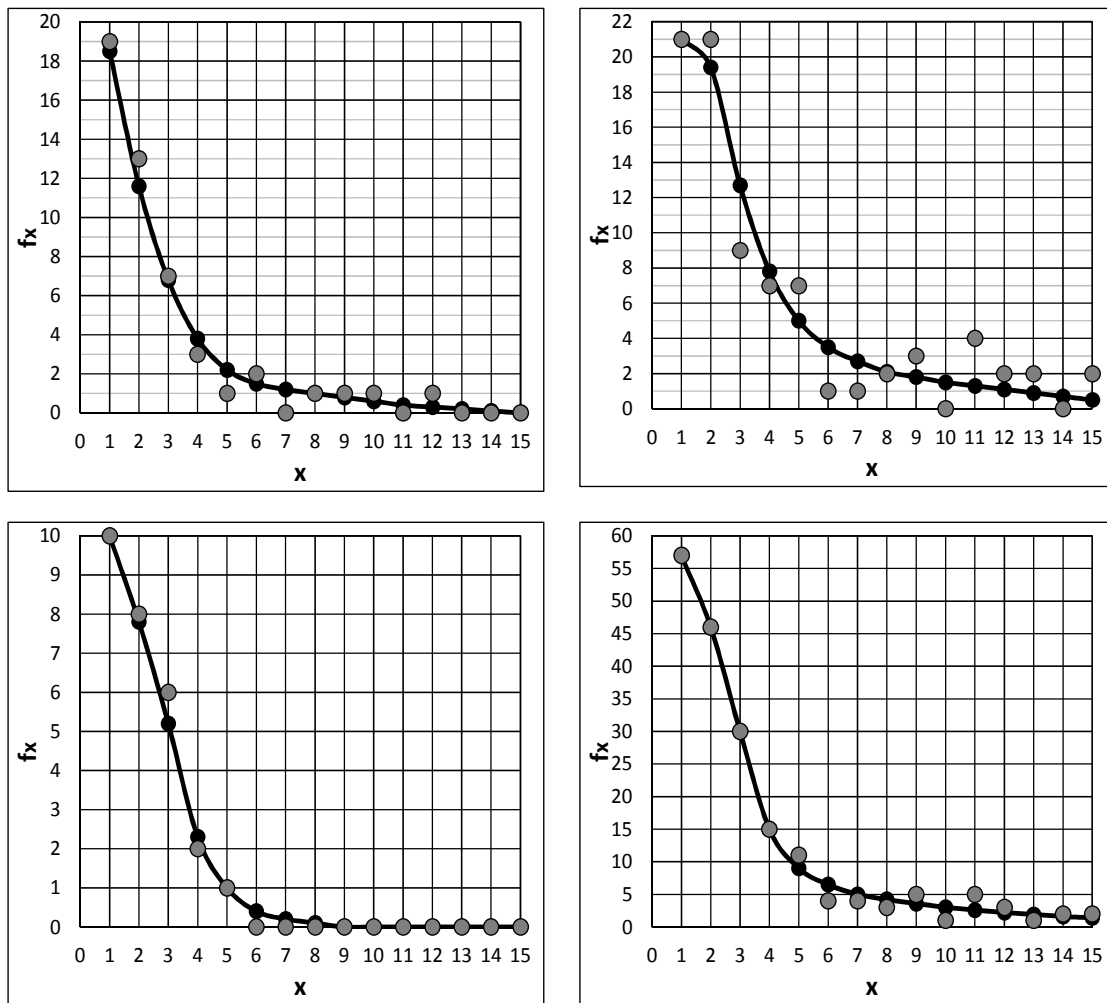
**N.B. 2:** In order to reduce the influence of drawing stochasticity on the values of the  $f_x$ , the as-recorded distribution of the  $f_x$  should preferably be smoothened: this may be obtained either by rarefaction processing or by regression of the as-recorded distribution of the  $f_x$  versus  $x$ .

**N.B. 3:** For  $f_1$  falling beneath  $0.6 \times f_2$  (that is when sampling completeness closely approaches exhaustivity), then Chao estimator may be selected: see reference [10].

## APPENDIX 2

**Regressions on the distributions of recorded values of  $f_x$  so as to reduce the consequences of drawing stochasticity**





**Figs. A2.1 to A2.6 – The recorded values of the numbers  $f_x$  of species recorded  $x$ -times (grey discs) and the regressed values of  $f_x$  (black discs) so as to reduce the consequence of stochastic dispersion, for the family *Papilionidae*, *Pieridae*, *Lycaenidae*, *Nymphalidae*, *Hesperidae* and for all five families together (from left to right and from top to bottom)**

© 2017 Béguinot; This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Peer-review history:  
The peer review history for this paper can be accessed here:  
<http://sciencedomain.org/review-history/19595>